Fuzzy Systems
Possibility Theory

Rudolf Kruse    Christian Moewes
{kruse,cmoewes}@iws.cs.uni-magdeburg.de

Otto-von-Guericke University of Magdeburg
Faculty of Computer Science
Department of Knowledge Processing and Language Engineering
Outline

1. Introduction
   Problems with Probability Theory
   Upper and Lower Probabilities

2. Dempster-Shafer Theory of Evidence

3. Possibility Theory
Problems with Probability Theory
Representation of Ignorance

• standard method to handle uncertainty is probability theory
• however, it has some problems regarding ignorance

we are given one die with faces 1, . . . , 6
what is the certainty of showing up face $i$?

1. conduct statistical survey (roll the die 10,000 times) and estimate relative frequency: $P(\{i\}) = \frac{1}{6}$
2. use subjective probabilities (i.e., often normal case): we don’t know anything (especially and explicitly we don’t have any reason to assign unequal probabilities), so most plausible distribution is uniform

⇒ problem: uniform distribution because of ignorance or extensive statistical tests
Problems with Probability Theory

Representation of Ignorance

• experts analyze aircraft shapes
  • 3 aircraft types A, B, C
  • “It is type A or B with 90% certainty.”
  • “About C, I don’t have any clue and I don’t want to commit myself.”
  • “No preferences for A or B.”

⇒ problem: propositions hard to handle with Bayesian theory
Upper and Lower Probabilities

- a way out to solve this problem are upper and lower probabilities
  - “It is type A or B with 90% certainty.”
  - “About C, I don’t have any clue and I don’t want to commit myself.”
  - “No preferences for A or B.”

<table>
<thead>
<tr>
<th>A</th>
<th>∅</th>
<th>{A}</th>
<th>{B}</th>
<th>{C}</th>
<th>{A, B}</th>
<th>{A, C}</th>
<th>{A, B, C}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_*(A)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$P^*(A)$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- consider $P_*(A)$ and $P^*(A)$ as lower and upper probability bounds
- total ignorance: $P_*(A) = 0$ and $P^*(A) = 1$ for $A \neq \emptyset, A \neq \Omega$
Properties of Upper and Lower Probabilities I

1. \( P^* : \mathcal{P}(\Omega) \to [0, 1] \)

2. \( 0 \leq P_* \leq P^* \leq 1, \quad P_*(\emptyset) = P^*(\emptyset) = 0, \quad P_*(\Omega) = P^*(\Omega) = 1 \)

3. \( A \subseteq B \implies P^*(A) \leq P^*(B) \) and \( P_*(A) \leq P_*(B) \)

4. \( A \cap B = \emptyset \not\implies P^*(A) + P^*(B) = P^*(A \cup B) \)

5. \( P_*(A \cup B) \geq P_*(A) + P_*(B) - P_*(A \cap B) \)

6. \( P^*(A \cup B) \leq P^*(A) + P^*(B) - P^*(A \cap B) \)

7. \( P_*(A) = 1 - P^*(\Omega \setminus A) \)
Properties of Upper and Lower Probabilities II

• one can prove the following generalized equation

\[ P^*(\bigcup_{i=1}^{n} A_i) \geq \sum_{\emptyset \neq I : I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \cdot P^*(\bigcap_{i \in I} A_i) \]

• these set functions play an important role in theoretical physics (i.e., capacities [Choquet, 1954])

• thoughts have been generalized and developed into the theory of belief functions [Shafer, 1976]
Outline

1. Introduction

2. Dempster-Shafer Theory of Evidence
   - Belief Measure
   - Plausibility Measure
   - Basic Belief Assignment
   - Focal Element
   - Total Ignorance
   - Dempster’s Rule of Combination
   - Example
   - Conversion Formulas

3. Possibility Theory
Dempster-Shafer Theory of Evidence

- like Bayesian approaches, the Dempster-Shafer theory of evidence [Shafer, 1976] wants to model and quantify uncertainty by degrees of belief
- Bayesian approaches assign degrees of belief to single hypothesis
- theory of evidence assigns degrees of belief to sets of hypotheses

- let \( X \) be finite “frame of discernment”
  \[ \Rightarrow \] each proposition “\( x \) is \( x_0 \)” can be represented by \( A \subset X \)
  - sets including one element correspond to elementary propositions
  - proposition is true if it contains the true elementary proposition
  - belief in “\( x \) is \( x_0 \)” related to \( A \subset X \) is measured by \( \text{Bel}(A) \in [0, 1] \)
Belief Measure

Definition

Given a finite universe $X$, a belief measure is a function

$$\text{Bel} : \mathcal{P}(X) \rightarrow [0, 1]$$

s.t. $\text{Bel}(\emptyset) = 0$, $\text{Bel}(X) = 1$ and

$$\text{Bel}(A_1 \cup A_2 \cup \ldots \cup A_n) \geq \sum_{i=1}^{n} \text{Bel}(A_j) - \sum_{i,j:1\leq i<j\leq n} \text{Bel}(A_i \cap A_j) + \ldots + (-1)^{n+1} \text{Bel}(A_1 \cap A_2 \cap \ldots \cap A_n)$$

for all possible families of subsets of $X$.

- e.g., $\text{Bel}(A_1 \cup A_2) \geq \text{Bel}(A_1) + \text{Bel}(A_2) - \text{Bel}(A_1 \cap A_2)$
- difficult to test $\Rightarrow$ basic belief assignment is introduced
Belief Measure

• for each $A \in \mathcal{P}(X)$, $\text{Bel}(A)$ is interpreted as **degree of belief** (based on evidence) that given $x \in X$ belongs to $A$

• if sets $A_1, \ldots, A_n$ are pairwise disjoint, then

$$\text{Bel}(A_1, \ldots, A_n) \geq \sum_{i=1}^{n} \text{Bel}(A_i)$$

$\Rightarrow$ weaker version of additivity property of probability measures

$\Rightarrow$ probability measures are special case of belief measures where equality in definition of $\text{Bel}$ is always satisfied

• fundamental property: let $n = 2$, $A_1 = A$ and $A_2 = A^C$, then

$$\text{Bel}(A) + \text{Bel}(A^C) \leq 1$$
Plausibility Measure

- associated with each Bel is **plausibility measure** \(\text{Pl}\) defined by

\[
\text{Pl}(A) = 1 - \text{Bel}(A^C), \quad \forall A \in \mathcal{P}(X)
\]

**Definition**

Given a finite universe \(X\), a **plausibility measure** is a function

\[
\text{Pl} : \mathcal{P} \rightarrow [0, 1]
\]

s.t. \(\text{Pl}(\emptyset) = 0\), \(\text{Pl}(X) = 1\) and

\[
\text{Pl}(A_1 \cap A_2 \cap \ldots \cap A_n) \leq \sum_j \text{Pl}(A_j) - \sum_{j<k} \text{Pl}(A_j \cup A_k)
\]

\[
\ldots + (-1)^{n+1} \text{Pl}(A_1 \cup A_2 \cup \ldots \cup A_n)
\]

for all possible families of subsets of \(X\).
Basic Belief Assignment

• fundamental property: let $n = 2$, $A_1 = A$ and $A_2 = A^C$, then
  \[ \text{Pl}(A) + \text{Pl}(A^C) \geq 1 \]

• Bel and Pl can be characterized by basic belief assignment
  \[ m : \mathcal{P}(X) \rightarrow [0, 1] \]
  s.t. $m(\emptyset) = 0$ and
  \[ \sum_{A \in \mathcal{P}(X)} m(A) = 1 \]

• for each $A \in \mathcal{P}(X)$, value $m(A)$ expresses proportion to which all evidence supports “particular element of $X$ belongs to $A$”

• $m(A)$ does not imply any claim regarding subsets of $A$

• definition of $m$ resembles probability distribution function, but
  • probability distribution functions are defined on $X$
  • basic belief assignments are defined on $\mathcal{P}(X)$
Basic Belief Assignment

- observe the following facts
  1. $m(X)$ need not to be 1
  2. if $A \subseteq B$, then $m(A)$ need not to be less or equal $m(B)$
  3. there does not need to be any relation between $m(A)$ and $m(A^C)$

- given $m$, Bel and Pl are uniquely determined for all $A \in \mathcal{P}(X)$ by

\[ \text{Bel}(A) = \sum_{B \mid B \subseteq A} m(B) \quad \text{and} \quad \text{Pl}(A) = \sum_{B \mid A \cap B \neq \emptyset} m(B) \]

- note that inverse is also possible
- e.g., $m(A) = \sum_{B \mid B \subseteq A} (-1)^{|A \setminus B|} \text{Bel}(B)$
- each of three functions is sufficient to determine other two
Basic Belief Assignment

• $\text{Bel}(A) = \sum_{B \mid B \subseteq A} m(B)$ means the following:
  • $m(A)$ characterizes degree of evidence that $x$ belongs only to $A$
  • $\text{Bel}(A)$ represents evidence that $x$ belongs to $A$ and subsets of $A$

• $\text{Pl}(A) = \sum_{B \mid A \cap B \neq \emptyset} m(B)$ represents both kinds of evidence:
  • total evidence that $x$ belongs to $A$ or to any subset of $A$
  • evidence associated with sets that overlap with $A$

$\Rightarrow$ for all $A \in \mathcal{P}(X)$,
$$\text{Pl}(A) \geq \text{Bel}(A)$$
Focal Element

- every $A \in \mathcal{P}(X)$ for which $m(A) > 0$ is called **focal element** of $m$
- they are subsets of $X$ on which available evidence focuses
- $X$ is finite $\Rightarrow m$ can be fully described by list $\mathcal{F}$ of its focal elements
- pair $\langle \mathcal{F}, m \rangle$ is called **body of evidence**
Total Ignorance

- **total ignorance** is expressed by

\[ m(X) = 1 \text{ and } m(A) = 0 \quad \forall A \neq X \]

- *i.e.*, we know that element is in universal set but we have no evidence about its location in any \( A \subset X \)

- in terms of belief it is exactly the same:

\[ Bel(X) = 1 \text{ and } Bel(A) = 0 \quad \forall A \neq X \]

- in terms of plausibility it is different:

\[ Pl(\emptyset) = 0 \text{ and } Pl(A) = 1 \quad \forall A \neq \emptyset \]
Dempster’s Rule of Combination

• now, consider two independent sources of evidence (e.g., from two experts) expressed by $m_1$ and $m_2$ on some $\mathcal{P}(X)$

• they must be appropriately combined to obtain joint basic assignment $m_{1,2}$

• generally, evidence can be combined in many ways

• standard way is expressed by Dempster’s rule of combination

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K}$$

for all $A \neq \emptyset$ and $m_{1,2}(\emptyset) = 0$ where

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$$
Dempster’s Rule of Combination

- \( m_1(B) \) (focusing on \( B \in \mathcal{P}(X) \)) and \( m_2(C) \) (focusing on \( C \in \mathcal{P}(X) \)) are combined by \( m_1(B) \cdot m_2(C) \) (focusing on \( B \cap C \))

\( \Rightarrow \) exactly the same for joint probability distribution which is calculated from two independent marginal distributions

- since some intersections of focal elements may result in \( A \), we must add corresponding product to obtain \( m_{1,2} \)
- also, some intersections may be empty
- since, \( m_{1,2}(\emptyset) = 0 \), \( K \) is thus not included in definition of \( m_{1,2} \)
- to ensure \( \sum_{A \in \mathcal{P}(X)} m(A) = 1 \), each product is divided by \( 1 - K \)
Example

- assume that an old code fragment was discovered by Rudolf Kruse
- it strongly resembles code fragments of Christian Moewes
- several questions can come up:
  1. Is the discovered code fragment a real fragment of Christian?
  2. Is the discovered code fragment a product of one of Christian’s students?
  3. Is the discovered code fragment a fake?
Example

- let $C, S, F$ denote subsets of universe $X$ of all code fragments
- $C$ denotes set of all code fragments by Christian
- $S$ represents set of all code fragments by Christian’s students
- $F$ represents set of all fakes of Christian’s code fragments

- two experts (named Georg and Matthias) performed careful examinations of the code fragment
- afterwards, they provided Rudolf Kruse with basic assignments $m_1$ and $m_2$ as follows
Two Independent Sources

<table>
<thead>
<tr>
<th>experts</th>
<th>Georg</th>
<th>Matthias</th>
</tr>
</thead>
<tbody>
<tr>
<td>focal elements</td>
<td>$m_1$</td>
<td>Bel$_1$</td>
</tr>
<tr>
<td>$C$</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>$S$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$F$</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>$C \cup S$</td>
<td>.15</td>
<td>.2</td>
</tr>
<tr>
<td>$C \cup F$</td>
<td>.1</td>
<td>.2</td>
</tr>
<tr>
<td>$S \cup F$</td>
<td>.05</td>
<td>.1</td>
</tr>
<tr>
<td>$C \cup S \cup F$</td>
<td>.6</td>
<td>1</td>
</tr>
</tbody>
</table>

- degrees of evidence obtained by examination and supported by different claims that code fragment belongs to one of the sets
- given $m_1$ and $m_2$, we can easily compute Bel$_1$ and Bel$_2$, resp.
Calculation of Normalization Factor

\[ K = m_1(C) \cdot m_2(S) + m_1(C) \cdot m_2(F) + m_1(C) \cdot m_2(S \cup F) \\
+ m_1(S) \cdot m_2(C) + m_1(S) \cdot m_2(F) + m_1(S) \cdot m_2(C \cup F) \\
+ m_1(F) \cdot m_2(C) + m_1(F) \cdot m_2(S) + m_1(F) \cdot m_2(C \cup S) \\
+ m_1(C \cup S) \cdot m_2(F) + m_1(C \cup F) \cdot m_2(S) + m_1(S \cup F) \cdot m_2(C) \]

\[ = .03 \]

• so, the normalization factor is then \( 1 - K = .97 \)
Calculation of Joint Basic Assignment

\[ m_{1,2}(C) = [m_1(C) \cdot m_2(C) + m_1(C) \cdot m_2(C \cup S) + m_1(C) \cdot m_2(C \cup F) + m_1(C \cup S) \cdot m_2(C) + m_1(C \cup S) \cdot m_2(C \cup S) + m_1(C \cup F) \cdot m_2(C) + m_1(C \cup F) \cdot m_2(C \cup S) + m_1(C \cup S \cup F) \cdot m_2(C)]/ .97 = .21 \]

\[ m_{1,2}(S) = [m_1(S) \cdot m_2(S) + m_1(S) \cdot m_2(C \cup S) + m_1(S) \cdot m_2(C \cup F) + m_1(S \cup C \cup S) \cdot m_2(S) + m_1(S \cup F) \cdot m_2(S) + m_1(S \cup F) \cdot m_2(C \cup S) + m_1(S \cup F) \cdot m_2(C \cup S) \cdot m_2(S)]/ .97 = .01 \]
Calculation of Joint Basic Assignment

\[
m_{1,2}(C \cup F) = [m_1(C \cup F) \cdot m_2(C \cup F) \\
+ m_1(C \cup F) \cdot m_2(C \cup S \cup F) \\
+ m_1(C \cup S \cup F) \cdot m_2(C \cup F)]/.97 \\
= .2
\]

\[
m_{1,2}(C \cup S \cup F) = [m_1(C \cup S \cup F) \cdot m_2(C \cup S \cup F)]/.97 \\
= .31
\]
Joint Basic Assignment

<table>
<thead>
<tr>
<th>focal elements</th>
<th>Georg</th>
<th>Matthias</th>
<th>combined evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m_1$</td>
<td>Bel$_1$</td>
<td>$m_2$</td>
</tr>
<tr>
<td>$C$</td>
<td>.05</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>$S$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$F$</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>$C \cup S$</td>
<td>.15</td>
<td>.2</td>
<td>.05</td>
</tr>
<tr>
<td>$C \cup F$</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
</tr>
<tr>
<td>$S \cup F$</td>
<td>.05</td>
<td>.1</td>
<td>.05</td>
</tr>
<tr>
<td>$C \cup S \cup F$</td>
<td>.6</td>
<td>1</td>
<td>.5</td>
</tr>
</tbody>
</table>
Projection

- consider now basic assignment $m: \mathcal{P}(X \times Y) \rightarrow [0, 1]$
- each focal element of $m$ is binary relation $R$ on $X \times Y$
- let $R_X$ denote projection of $R$ on $X$, then

$$R_X = \{x \in X \mid (x, y) \in R \text{ for some } y \in Y\}$$

- projection $m_X$ of $m$ on $X$ is

$$m_X(A) = \sum_{R \mid A = R_X} m(R) \text{ for all } A \in \mathcal{P}(X)$$

- adding values of $m(R)$ for all focal elements $R$ whose $R_X$ is $A$
- similarly, computation can be done for $R_Y$ and $m_Y$, resp.
Marginal Bodies of Evidence

• $m_X$ and $m_Y$ are also called marginal basic assignments

• $\langle F_X, m_X \rangle$ and $\langle F_Y, m_Y \rangle$ are called marginal bodies of evidence

• $\langle F_X, m_X \rangle$ and $\langle F_Y, m_Y \rangle$ are called noninteractive if and only if for all $A \in F_X$ and $B \in F_Y$

$$m(A \times B) = m_X(A) \cdot m_Y(B)$$

and

$$m(R) = 0 \text{ for all } R \neq A \times B$$

• i.e., joint basic assignment $m$ is uniquely determined by marginal basic assignments
Example

- consider body of evidence given in table
- focal elements are subsets of $X \times Y = \{1, 2, 3\} \times \{a, b, c\}$

<table>
<thead>
<tr>
<th>focal element</th>
<th>$X \times Y$</th>
<th>$m(R_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1 =$</td>
<td>0 0 0 0 1 1 0 1 1</td>
<td>.0625</td>
</tr>
<tr>
<td>$R_2 =$</td>
<td>0 0 0 1 0 0 1 0 0</td>
<td>.0625</td>
</tr>
<tr>
<td>$R_3 =$</td>
<td>0 0 0 1 1 1 1 1 1</td>
<td>.125</td>
</tr>
<tr>
<td>$R_4 =$</td>
<td>0 1 1 0 0 0 1 1</td>
<td>.0375</td>
</tr>
<tr>
<td>$R_5 =$</td>
<td>0 1 1 0 1 1 0 0 0</td>
<td>.075</td>
</tr>
<tr>
<td>$R_6 =$</td>
<td>0 1 1 0 1 1 1 1</td>
<td>.075</td>
</tr>
<tr>
<td>$R_7 =$</td>
<td>1 0 0 0 0 0 1 0 0</td>
<td>.375</td>
</tr>
<tr>
<td>$R_8 =$</td>
<td>1 0 0 1 0 0 0 0 0</td>
<td>.075</td>
</tr>
<tr>
<td>$R_9 =$</td>
<td>1 0 0 1 0 0 1 0 0</td>
<td>.075</td>
</tr>
<tr>
<td>$R_{10} =$</td>
<td>1 1 1 0 0 0 1 1 1</td>
<td>.075</td>
</tr>
<tr>
<td>$R_{11} =$</td>
<td>1 1 1 1 1 1 0 0 0</td>
<td>.15</td>
</tr>
<tr>
<td>$R_{12} =$</td>
<td>1 1 1 1 1 1 1 1</td>
<td>.15</td>
</tr>
</tbody>
</table>
Example

- margin bodies of evidence are obtained as follows

\[ m_X(\{2, 3\}) = m(R_1) + m(R_2) + m(R_3) = .25 \]
\[ m_X(\{1, 2\}) = m(R_5) + m(R_8) + m(R_{11}) = .3 \]
\[ m_Y(\{a\}) = m(R_2) + m(R_7) + m(R_8) + m(R_9) = .25 \]
\[ m_Y(\{a, b, c\}) = m(R_4) + m(R_{10}) + m(R_{11}) + m(R_{12}) = .5 \]

<table>
<thead>
<tr>
<th>A =</th>
<th>0 1 1</th>
<th>m_X(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 1</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>1 1 0</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>1 1 1</td>
<td>.3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B =</th>
<th>0 1 1</th>
<th>m_Y(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 0</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>1 1 1</td>
<td>.5</td>
<td></td>
</tr>
</tbody>
</table>

- here, marginal bodies of evidence are noninteractive
Conversion Formulas

<table>
<thead>
<tr>
<th>given</th>
<th>( m(A) = )</th>
<th>( \text{Bel}(A) = )</th>
<th>( \text{Pl}(A) = )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>( m(A) )</td>
<td>( \sum_{B \subseteq A} m(B) )</td>
<td>( \sum_{B \cap A \neq \emptyset} m(B) )</td>
</tr>
<tr>
<td>( \text{Bel} )</td>
<td>( \sum_{B \subseteq A} (-1)^{</td>
<td>A \setminus B</td>
<td>} \text{Bel}(B) )</td>
</tr>
<tr>
<td>( \text{Pl} )</td>
<td>( \sum_{B \subseteq A} (-1)^{</td>
<td>A \setminus B</td>
<td>} (1 - \text{Pl}(B^C)) )</td>
</tr>
</tbody>
</table>
Outline

1. Introduction

2. Dempster-Shafer Theory of Evidence

3. Possibility Theory
   - Necessity and Possibility
   - Possibility Distribution
   - Basic Distribution
   - Mathematical Properties
   - Fuzzy Sets and Possibility Theory
   - Possibility versus Probability
   - Comparison of both Theories
   - Interpretations
Possibility Theory

- **possibility theory** is a special kind of evidence theory.
- Bodies of evidence contain **nested** focal elements.
- Belief and plausibility are called **necessity** and **possibility**, respectively.

**Theorem**

Let a given finite body of evidence $\langle \mathcal{F}, m \rangle$ be nested. Then, the associated belief and plausibility measures have the following properties for all $A, B \in \mathcal{P}(X)$

$$\text{Bel}(A \cap B) = \min\{\text{Bel}(A), \text{Bel}(B)\}$$
$$\text{Pl}(A \cup B) = \max\{\text{Pl}(A), \text{Pl}(B)\}$$
Necessity and Possibility

- let necessity and plausibility be denoted by Nec and Pos, resp.
- last theorem provides us with basic equation of possibility theory

\[
\text{Nec}(A \cap B) = \min\{\text{Nec}(A), \text{Nec}(B)\}
\]
\[
\text{Pos}(A \cup B) = \max\{\text{Pos}(A), \text{Pos}(B)\}
\]

- can be also defined for arbitrary universal set (e.g., infinite ones)

\[
\text{Nec}\left(\bigcap_{k \in K} A_k\right) = \inf_{k \in K} \text{Nec}(A_k)
\]
\[
\text{Pos}\left(\bigcup_{k \in K} A_k\right) = \sup_{k \in K} \text{Pos}(A_k)
\]

where \( K \) is arbitrary index set
Necessity and Possibility

- since necessity and possibility are special belief and plausibility measures, resp., they satisfy fundamental properties

\[ \text{Nec}(A) + \text{Nec}(A^C) \leq 1 \]
\[ \text{Pos}(A) + \text{Pos}(A^C) \geq 1 \]
\[ 1 - \text{Pos}(A^C) = \text{Nec}(A) \]

- from the definitions of Nec and Pos, it follows that

\[ \min \left\{ \text{Nec}(A), \text{Nec}(A^C) \right\} = 0 \]
\[ \max \left\{ \text{Pos}(A), \text{Pos}(A^C) \right\} = 1 \]
Necessity and Possibility

- necessity and possibility measures constrain each other

**Theorem**

*For every $A \in \mathcal{P}(X)$, any necessity measure $\text{Nec}$ on $\mathcal{P}(X)$ and the associated possibility measure $\text{Pos}$ satisfy the following implications:*

\[
\text{Nec}(A) > 0 \Rightarrow \text{Pos}(A) = 1, \\
\text{Pos}(A) < 1 \Rightarrow \text{Nec}(A) = 0.
\]
Possibility Distribution Function

- given possibility measure $\text{Pos}$ on $\mathcal{P}(X)$
- let function $r : X \to [0, 1]$ s.t. $\forall x \in X : r(x) = \text{Pos}(\{x\})$ be called possibility distribution function associated with $\text{Pos}$

**Theorem**

*Every possibility measure $\text{Pos}$ on a finite power set $\mathcal{P}(X)$ is uniquely determined by a possibility distribution function*

\[ r : X \to [0, 1] \]

*via the formula*

\[ \text{Pos}(A) = \max_{x \in A} r(x) \]

*for each $A \in \mathcal{P}(X)$.*

- when $X$ is infinite, then $\text{Pos}(A) = \sup_{x \in A} r(x)$
• let possibility distribution function $r$ be defined on $X = \{x_1, \ldots, x_n\}$

$\Rightarrow$ **possibility distribution** associated with $r$ is $n$-tuple

$$r = (r_1, \ldots, r_n)$$

where $r_i = r(x_i)$ for all $x_i \in X$

• it is useful to order $r$ s.t. $r_i \geq r_j$ when $i < j$
Example

- now, let Pos be defined on $\mathcal{P}(X)$ by $m$
- \textit{w.l.o.g.}, focal elements are some or all of subsets in complete sequence of nested subsets

$$A_1 \subset A_2 \subset \ldots \subset A_n = X$$

where $A_i = \{x_1, \ldots, x_i\}$ $i \in \mathbb{IN}$

- \textit{i.e.}, $m(A) = 0$ for each $A \neq A_i$ ($i \in \mathbb{IN}$) and $\sum_{i=1}^{n} m(A_i) = 1$
Example

- complete sequence of nested focal elements of Pos on $\mathcal{P}(X)$
  where $X = \{x_1, \ldots, x_n\}$
Basic Distribution

• every Pos on finite set can be represented by $n$-tuple

$$m = (m_1, \ldots, m_n)$$

for some finite $n \in \mathbb{N}$ where $m_i = m(A_i)$ for all $i \in \mathbb{N}_n$

• clearly, $\sum_{i=1}^n m_i = 1$ and $m_i \in [0, 1]$ for all $i \in \mathbb{N}_n$

• $m$ is called basic distribution

• each $m$ represents exactly one possibility distribution $r$

$$\forall x_i \in X : r_i = r(x_i) = \text{Pos}({x_i}) = \text{Pl}({x_i})$$

$$\forall i \in \mathbb{N}_n : r_i = \text{Pl}({x_i}) = \sum_{k=i}^{n} m(A_k) = \sum_{k=i}^{n} m_k$$

• solving all $n$ equations for $m_i$ ($i \in \mathbb{N}_n$) leads to

$$m_i = r_i - r_{i+1}$$

• one-two-one correspondence between $r$ and $m$
Example

- one possibility measure defined on $X$ with $A_1 \subset A_2 \subset \ldots \subset A_7$

\[ m(A_6) = .1 \]
\[ m(A_3) = .4 \]
\[ m(A_7) = .2 \]

\[ m(A_2) = .3 \]
\[ r(x_1) = 1 \]
\[ r(x_2) = 1 \]
\[ r(x_3) = .7 \]
\[ r(x_4) = .3 \]
\[ r(x_5) = .3 \]
\[ r(x_6) = .3 \]
\[ r(x_7) = .2 \]

\[ \sum_{i=1}^{n} m(A_i) = 1 \text{ is satisfied}, \ m(A_1) = m(A_4) = m(A_5) = 0 \]
Perfect Evidence and Total Ignorance

- smallest possibility distribution \( \tilde{r} \) has form

\[
\tilde{r} = (1, 0, 0, \ldots, 0)
\]

- \( \tilde{r} \) represents perfect evidence with no uncertainty involved

- largest possibility distribution \( \hat{r} \) can be written as

\[
\hat{r} = (0, 0, \ldots, 0, 1)
\]

- \( \hat{r} \) denotes total ignorance, i.e., no relevant evidence is available

- the larger \( r \), the less specific the evidence is (the more ignorant)
Joint Possibility Distribution

• consider joint possibility distributions \( r \) defined on \( X \times Y \)

• projections \( r_X, r_Y \) are called marginal possibility distribution

\[
\forall x \in X : r_X(x) = \max_{y \in Y} r(x, y)
\]

\[
\forall y \in Y : r_Y(y) = \max_{x \in X} r(x, y)
\]

• formula for \( r_X(x) \) follows from

\[
\text{Pos}_X(\{x\}) = \text{Pos}(\{(x, y) \mid x \in Y\})
\]

\[
\text{Pos}_X(\{x\}) = r_X(x)
\]

\[
\text{Pos}(\{(x, y) \mid x \in Y\}) = \max_{y \in Y} r(x, y)
\]

• same argumentation can be done for \( r_Y \)
Possibilistic Noninteraction

Definition

Nested bodies of evidence $X$ and $Y$ represented by $r_X$ and $r_Y$, resp., are called noninteractive if and only if

$$r(x, y) = \min\{r_X(x), r_Y(y)\}$$

for all $x \in X$ and all $y \in Y$

- recall product rule $m(A \times B) = m_X(A) \cdot m_Y(B)$
- this definition is not based upon product rule of evidence theory
  \(\Rightarrow\) does not conform to general noninteractive bodies of evidence
  - here, product rule would not preserve nested structure
Possibilistic Noninteraction

- let $r_X$ and $r_Y$ be marginal poss. dist. functions on $X$ and $Y$, resp.
- let $r$ be joint poss. dist. function on $X \times Y$ defined by $r_X$ and $r_Y$
- assume $\text{Pos}_X$, $\text{Pos}_Y$ and $\text{Pos}$ correspond to $r_X$, $r_Y$ and $r$, then

$$\text{Pos}(A \times B) = \min\{\text{Pos}_X(A), \text{Pos}_Y(B)\}$$

for all $A \in \mathcal{P}(X)$ and all $B \in \mathcal{P}(Y)$ where

$$\text{Pos}_X(A) = \max_{x \in A} r_X(x),$$

$$\text{Pos}_Y(B) = \max_{y \in B} r_Y(y),$$

$$\text{Pos}(A \times B) = \max_{x \in A, y \in B} r(x, y)$$
Possibilistic Independence

- two marginal possibilistic bodies of evidence are **independent** if and only if conditional possibilities equal marginal possibilities, *i.e.*, for all $x \in X$ and all $y \in Y$

  $$r_{X|Y}(x|y) = r_X(x),$$
  $$r_{Y|X}(y|x) = r_Y(y)$$

- $r_{X|Y}(x|y)$ and $r_{Y|X}(y|x)$ are conditional possibilities on $X \times Y$

- possibilistic independence is stronger than poss. noninteraction
- possibilistic independence $\Rightarrow$ possibilistic noninteraction
- possibilistic noninteraction $\nRightarrow$ possibilistic independence
Fuzzy Sets and Possibility Theory

- possibility theory can be also formulated using fuzzy sets instead of nested bodies of evidence [Dubois and Prade, 1993]
- fuzzy sets (similar to bodies of evidence) are based on families of nested sets, i.e., $\alpha$-cuts
- Pos measures are connected with fuzzy sets via possibility distribution function

- e.g., let $X$ denote variable taking values in set $X$
- $X = x$ describes “$X$ is $x$” where $x \in X$
- fuzzy set $F$ on $X$ expresses constraint on values assigned to $X$
- given value $x \in X$, $F(x)$ is degree of compatibility of $x$ described by $F$
Fuzzy Sets and Possibility Theory

• given “$\mathcal{X}$ is $F$”, $F(x)$ is **degree of possibility** that $\mathcal{X} = x$

• *i.e.*, given $F$ on $X$ and “$\mathcal{X}$ is $F$”,

\[ \forall x \in X : r_F(x) = F(x) \]

• given $r_F$, associated possibility measure is

\[ \forall A \in \mathcal{P}(X) : Pos_F(A) = \sup_{x \in A} r_F(x) \]

• for normal fuzzy sets, $\text{Nec}_F$ can be calculated by

\[ \forall A \in \mathcal{P}(X) : \text{Nec}_F(A) = 1 - Pos_F(A^C) \]
Example

- let variable $T$ be temperature in °C (only integers)
- actual value is given in terms of "$T$ is around 21°C"

\[ F = r_F \]

- incomplete information induces possibility distribution function $r_F$
- $r_F$ is numerically identical with membership function $F$
- nested $\alpha$-cuts play same role as focal elements
- **Question**: Which values do Nec and $m$ take here?
Summary

• if $r_F$ is derived from normal fuzzy set $F$, then both formulations of possibility theory are equivalent
• full equivalence breaks when $F$ is not normal
  $\Rightarrow$ basic probability assignment function $m$ is not directly applicable
• all other properties remain equivalent in both formulations

• possibility theory is measure-theoretic counterpart of fuzzy set theory based upon standard fuzzy operations
• it provides tools for processing incomplete information expressed by fuzzy propositions
• it plays major role in fuzzy logic and approximate reasoning
Possibility versus Probability

• probability theory and possibility theory are distinct
• to show that, let us examine both theories by evidence theory
• probability theory and possibility theory are special branches of it
⇒ introduce probabilistic counterparts of possibilistic characteristics

• probability measure Pro must satisfy

\[ \text{Pro}(A \cup B) = \text{Pro}(A) + \text{Pro}(B) \]

for all \( A, B \in \mathcal{P}(X) \) s.t. \( A \cap B = \emptyset \)
• Pro is special type of belief measure
Probability Distribution Function

**Theorem**

A belief measure $\text{Bel}$ on a finite power set $\mathcal{P}(X)$ is a probability measure if and only if the associated basic probability assignment function $m$ is given by $m(\{x\}) = \text{Bel}(\{x\})$ and $m(A) = 0$ for all subsets of $X$ that are no singletons.

$\Rightarrow$ probability measures on finite sets are thus fully represented by

$$p : X \rightarrow [0, 1]$$

s.t. $p(x) = m(\{x\})$

- $p$ is called **probability distribution function**
- let $p = (p(x) \mid x \in X)$ be **probability distribution** on $X$
Probability Measure

• recall definition of belief and plausibility

\[
\text{Bel}(A) = \sum_{B \mid B \subseteq A} m(B) \quad \text{and} \quad \text{Pl}(A) = \sum_{B \mid A \cap B \neq \emptyset} m(B)
\]

• if basic probability assignment focuses only on singletons, then

\[
\forall A \in \mathcal{P}(X) : \text{Bel}(A) = \text{Pl}(A) = \sum_{x \in A} m(\{x\})
\]

• Bel and Pos are equal \( \Rightarrow \) it is convenient to use one symbol Pro

\[
\forall A \in \mathcal{P}(X) : \text{Pro}(A) = \sum_{x \in A} p(x)
\]

• Pro is called **probability measure**
## Comparison of both Theories

<table>
<thead>
<tr>
<th>Property</th>
<th>Probability Theory</th>
<th>Possibility Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>measures</td>
<td>probability Pro</td>
<td>possibility Pos, necessity Nec</td>
</tr>
<tr>
<td>body of evidence</td>
<td>singletons</td>
<td>family of nested subsets</td>
</tr>
<tr>
<td>representation</td>
<td>$p : X \rightarrow [0, 1]$ via $\text{Pro}(A) = \sum_{x \in A} p(x)$</td>
<td>$r : X \rightarrow [0, 1]$ via $\text{Pos}(A) = \max_{x \in A} r(x)$</td>
</tr>
<tr>
<td>normalization</td>
<td>$\sum_{x \in X} p(x) = 1$</td>
<td>$\max_{x \in X} r(x) = 1$</td>
</tr>
<tr>
<td>total ignorance</td>
<td>$\forall x \in X : p(x) = \frac{1}{</td>
<td>X</td>
</tr>
<tr>
<td>noninteraction</td>
<td>$p(x, y) = p_X(x) \cdot p_Y(y)$</td>
<td>$r(x, y) = \min{r_X(x), r_Y(y)}$</td>
</tr>
<tr>
<td>independence</td>
<td>$p_{X</td>
<td>Y}(x</td>
</tr>
<tr>
<td></td>
<td>$p_{Y</td>
<td>X}(y</td>
</tr>
<tr>
<td>independence $\iff$ noninteraction</td>
<td>independence $\Rightarrow$ noninteraction</td>
<td></td>
</tr>
</tbody>
</table>
### Comparison of both Theories

<table>
<thead>
<tr>
<th>Probability Theory</th>
<th>Possibility Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Pro}(A \cup B) = \text{Pro}(A) + \text{Pro}(B) - \text{Pro}(A \cap B) )</td>
<td>( \text{Pos}(A \cup B) = \text{max}{\text{Pos}(A), \text{Pos}(B)} )</td>
</tr>
<tr>
<td></td>
<td>( \text{Nec}(A \cap B) = \text{min}{\text{Nec}(A), \text{Nec}(B)} )</td>
</tr>
<tr>
<td>not applicable</td>
<td>( \text{Nec}(A) = 1 - \text{Pos}(A^C) )</td>
</tr>
<tr>
<td></td>
<td>( \text{Pos}(A) &lt; 1 \Rightarrow \text{Nec}(A) = 0 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Nec}(A) &gt; 0 \Rightarrow \text{Pos}(A) = 1 )</td>
</tr>
<tr>
<td>( \text{Pro}(A) + \text{Pro}(A^C) = 1 )</td>
<td>( \text{Pos}(A) + \text{Pos}(A^C) \geq 1 )</td>
</tr>
<tr>
<td></td>
<td>( \text{Nec}(A) + \text{Nec}(A^C) \leq 1 )</td>
</tr>
<tr>
<td></td>
<td>( \text{max}{\text{Pos}(A), \text{Pos}(A^C)} = 1 )</td>
</tr>
<tr>
<td></td>
<td>( \text{min}{\text{Nec}(A), \text{Nec}(A^C)} = 0 )</td>
</tr>
</tbody>
</table>
FUZZY MEASURES: monotonic and continuous or semi continuous

BELIEF MEASURES: superadditive and continuous from above

PROBABILITY MEASURES: additive

PLAUSIBILITY MEASURES: subadditive and continuous from below

NEC. MEASURES

POS. MEASURES
Discussion

- possibility, necessity and probability measures do not overlap
- except for one special measure characterized by only one focal element (i.e., singleton)

⇒ distribution functions of possibility and probability become equal
- one element of universal set is assigned 1, all others assigned 0

⇒ this measure represents perfect evidence
Discussion

• differences in mathematical properties make each theory suitable for modeling certain types of uncertainty

• probability theory is ideal for formalizing uncertainty in situations
  • where class frequencies are known or
  • evidence is based on outcomes of many independent random experiments

• possibility theory is ideal for formalizing incomplete information expressed by fuzzy propositions

• necessity and possibility can be seen as lower and upper probabilities [Dempster, 1967]
Interpretations of Possibility Theory

based on similarity

- possibility \( r(x) \) reflects degree of similarity between \( x \) and prototype \( x_i \) having \( r(x_i) = 1 \)
- \( i.e. \), \( r(x) \) is expressed by distance between \( x \) and \( x_i \)
- the closer \( x \) to \( x_i \), the more possible its interpretation
- closeness comes from measurement or subjective judgement

based on preference

- due to ordering \( \leq_{\text{Pos}} \) defined on \( \mathcal{P}(X) \)
- \( \forall A, B \in \mathcal{P}(X), A \leq_{\text{Pos}} B \) means \( B \) is at least as possible as \( A \)
- \( A \leq_{\text{Pos}} B \iff A^C \leq_{\text{Nec}} B^C \)
Literature for the Lecture


